Deriving a Study Design to Examine the Frame of Reference Effect in Faking Instructions

Masterarbeit
im
Masterstudiengang Psychologie
der
Otto-Friedrich-Universität Bamberg

Verfasserin:
Mia Degro

Prüferin:
Prof. Dr. Astrid Schütz
Lehrstuhl für Persönlichkeitspsychologie
und Psychologische Diagnostik

DOI: https://doi.org/10.20378/irb-108706
URN: urn:nbn:de:bvb:473-irb-108706x

Deriving a Study Design to Examine the Frame of Reference Effect in Faking Instructions

Abstract

To study faking as an intentional response distortion in laboratory setting, scientists manipulate the generic instructions of measurement instruments such as personality questionnaires. Therefore, researchers generated applicant settings to elicit a motivation to fake and an as natural as possible faking behavior. Participants are instructed to answer the items as if they were applying for a job they really wanted and as if they wanted to maximize their chances of getting this job. The emerging “honest” and “faking” conditions are then compared to each other, and the difference is interpreted to result from faking.

Interestingly the frame of reference (FOR) research found the contextualization via instructions, which ask participants to think of their behavior in a certain context, e.g., at work while answering the items, to lead to a higher validity of contextualized questionnaires as compared to noncontextualized ones.

We argue that applicant faking instructions to study faking in laboratory settings do not only induce faking but at the same time a specific frame of reference. Thus, we assume that faking and frame of reference effects are confounded if this paradigm is used to study faking. If this holds to be true, studies using this manipulation of instructions do not meet the criterion of internal validity. If two systematically confounded processes cause the observed effects, the conclusion that these effects are caused by faking alone do not take the dependence of both factors into account and may therefore be biased.

Deriving a Study Design to Examine the Frame of Reference Effect in Faking Instructions

The body of research dealing with faking behavior is vast and very heterogenous. There is qualitative (e.g., König et al. 2012) and quantitative (e.g., Schulze et al., 2021) laboratory (e.g., Tett et al., 2012) and field (e.g., Detrick & Roberts, 2021) research regarding faking good (e.g., Williams et al., 2019) and faking bad (e.g., Bensch et al., 2019; Geiger et al., 2021). Faking is studied in different contexts (e.g., personnel selection: Goffin & Woods, 1995; van Hooft & Born, 2012; symptoms of psychological or physiological disorders: DePage & Merckelbach, 2021) and with different methods (i.e., interviews: e.g., Bourdage et al., 2020), questionnaires: e.g., Paulhus, 2011, indirect measures: Röhner et al., 2011). There are studies and models trying to understand the antecedents (e.g., Bill et al., 2020), the process (e.g., Roulin et al., 2016), influencing factors (e.g., McFarland & Ryan, 2000) and the consequences of faking (e.g., Niessen et al., 2017). Studies also concern different types of faking (e.g., McLarnon et al., 2019), and ways to deal with faking by preventing it (e.g., Wetzel et al., 2020), detecting it (e.g., Calanna et al., 2020), or correcting for the effects of intended response distortion (Reeder & Ryan, 2011).

Still, there is no uniform definition of faking, but there are some propositions which may serve as the least common denominator: Faking is an intentional, goal oriented and (socially) motivated behavior, emerging from an interaction between personal and situational variables to create a deceptive and inaccurate impression (Röhner & Schütz, 2020). It is assumed by most faking models that motivation, ability, and opportunity to fake are necessary for this behavior to occur (e.g., Ellingson & McFarland, 2011; Roulin et al., 2016).

But not only the research interests in this field are abundant. Also, the conclusions and findings vary and sometimes even contradict each other. For example, some researchers refer to the negative consequences of faking behavior in real life scenarios when decisions are based on diagnostic instruments like questionnaires or interviews, e.g., changes in rank order of applicants (Birkeland et al., 2006). But there are also others who argue that faking is associated with desired personal characteristics, such as social-emotional abilities (Geiger et al., 2021; Marcus, 2009) and may therefore include desirable information. There seems to be common ground regarding the following assumptions: Faking is possible, and individuals do fake their answers in different diagnostic procedures (e.g., IATs: Röhner et al. 2011; Interviews: Melchers et al., 2020, Questionnaires: Martin et al., 2002). Faking can have consequences regarding statistical characteristics (e.g., factor structure, means) and can thus influence consecutive decisions (Jeong et al., 2017; Schmitt & Bradburn, 2018; Sisco, 2010).

Despite this, the antecedents, the consequences, and especially and fundamentally the process of faking behavior are not fully understood (Ziegler et al., 2012). There are studies which found influencing factors, e.g., general mental abilities (e.g., Geiger, 2018), personality (see Heggestad, 2012), the job which the answers are faked for (e.g., Birkeland et al., 2006), or the absence or presence of warnings (e.g., Law et al., 2016). Still, research is needed to identify the causes of the heterogeneity in findings and to close the existing research gaps.

The challenge when studying faking is that this behavior is not socially desired but associated with negative emotions such as guilt (Li et al., 2021). Thus, those who fake are not likely to tell researchers that they did so. Consequently, especially with regard to naturally occurring faking behavior, faking has to be detected somehow. This proved to be very difficult (Ziegler & Bensch, 2015). One common way to do so is the usage of so-called lie scales, but these scales are often criticized (e.g., Stewart et al., 2010). Hence, field studies using faking detection methods are difficult to interpret (Röhner et al., 2022).

That is why probably the more precise way to study faking is to systematically induce this behavior in the laboratory. Mean differences between (quasi-)experimental honest and faking groups are within this paradigm used to estimate the degree of faking (e.g., Bensch et al., 2019). Laboratory studies that compare groups which are either instructed to fake or to answer honestly are an important avenue to gain insights into faking strategies.

Faking in Laboratory Research

To induce faking in the laboratory setting, scientists manipulate the instructions of questionnaires. Generic instructions ask participants to answer spontaneously or honestly, describe their typical behavior, or remind the respondents that there are no right or wrong answers and that they should choose the answers that fits best for them (Birkeland et al., 2006; Hartman & Lee Grub, 2011; Salgado & Lado, 2018). To induce faking, there are some commonly used manipulations which vary in their directness.

Some instructions are rather blunt and simply ask participants to fake their answers and state that honesty is not relevant (Davison et al., 2021, McFarland & Ryan, 2000). These instructions are often criticized as they do elicit the faking behavior but not the necessary motivation to fake (Mueller-Hansen et al., 2003) – but such motivation is considered an important antecedent of natural occurring faking behavior in most accounts (e.g., Roulin et al., 2016). Instead, such instructions induce an unnatural kind of faking which may be considered the upper boundary of a person's faking ability (Davison et al., 2021; LeBreton et al., 2007; Viswesveran & Ones, 1999).

Alternatively, more indirect faking instructions are used which are supposed to resemble real-life scenarios: To elicit a motivation to fake and a more natural behavior, researchers generated settings in which participants imagine to be applicants. They are instructed to answer the items as if they were applying for a job that they really want and to act in a way to maximize their chances of getting hired (e.g., Davison et al., 2021; Harris et al., 2020). The emerging "honest" and "faking" conditions are then compared to each other, and the difference is interpreted as faking (e.g., Bensch et al., 2019; Griffith et al., 2007; Tonković et al., 2012).

This conclusion seems straightforward. Laboratory studies have high internal validity as emerging group differences can be attributed to the experimentally manipulated variable, i.e., the instructions (Wirtz, 2021). Other systematic effects can be ruled out by the randomized assignment to the groups.

Still, we argue that there may be a problem that undermines the internal validity. It is possible that negative effects of faking behavior are underestimated because the manipulation may not only induce faking but may also include frame of reference effects (see also Anglim et al., 2021; Bing et al., 2004) – and that the two effects are intertwined. To separate the effects and know to which degree faking is due to the instruction per se, frame of reference effects would need to be controlled.

Frame of Reference

Research on frame of reference effects is grounded in an interactionist perspective which posits that personal and situational variables interact with each other e.g., by the CAPS-Theory (Mischel & Shoda, 1995). Following this rationale, the congruence between collected data and predicted behavior has to be increased to maximize the practical benefit of diagnostic instruments. If behavioral tendencies within a certain context are of interest, diagnostic instruments should cover this context (Lievens et al., 2008; Swift & Peterson, 2019). This conclusion is examined and supported by the frame of reference research.

It is assumed that participants draw a mental sample of their own behavior when answering to the items of a questionnaire (König et al., 2012; Kunda & Sanitioso, 1989; Lievens et al., 2008). While doing this, some of them may draw this sample across different situations while others may think of their behavior within a specific context such as their workplace – they adopt a certain frame of reference while answering to the items. There is between-person variability regarding these adopted frames of reference, meaning that some participants might have the tendency to think more about their workplace, while others rather think about the private context (König et al., 2012; Lievens et al., 2008; Robie et al., 2000).

Also, there is a within-person inconsistency as individuals do not answer every item with the same adopted frame of reference (König et al., 2012; Lievens et al., 2008). The context individuals think of while answering items is among individual differences (Robie et al., 2000) influenced by the item itself. Schulze et al. (2021) showed that even a supposedly generic item can provoke an inherent or "hidden" frame of reference depending on the subject they cover. For example, the visit of an art-gallery or a party (openness) is more of a private activity than a job-related one. On the other hand, leadership emergence or the conscientious completion of tasks (conscientiousness) is rather typical for work-related contexts. As the hidden frames of reference vary from item to item, they may contribute to the within-person inconsistency of adopted frames of reference and may also – if the hidden frame of reference is incongruent with the criterion – limit the overlap between items and criterion.

To limit the mental search space consistently across both participants and items, the whole questionnaire can be uniformly contextualized (e.g., Lievens et al., 2008). To create such a contextualization, either the instructions are manipulated (e.g., "Describe yourself as you typically are at work.", Shaffer & Postlethwaite, 2012, p. 484), or context tags are added to each item (i.e., item-level contextualization, e.g., I am rather quiet *at work*.). Thus, participants are given a certain frame of reference to adopt (e.g., at work, at school, or at home) while recalling their own behavior to answer the items of the questionnaire.

Researchers then compare the outcomes of the generic questionnaire with generic items (e.g., "I am rather quiet.", (Danner et al., 2016)) and instructions (e.g., "Read each statement and choose the one that best describes you.", Shaffer & Postlethwaite, 2012, p. 484) with a contextualized condition to study frame of reference effects.

For example, Lievens et al. (2008) compared an "at school" with an "at work" frame of reference and showed that contextualization makes a difference by reducing within-person inconsistency and between-person variability of adopted frames of reference. They also found that the frame of reference hast to be appropriately contextualized i.e., congruent with the criterion to take its positive effect. An "at work" frame of reference did not increase validity within the school context (i.e., predicting GPA) but an "at school" frame of reference did.

In summary, frame of reference effects are partly inherent in items from personality questionnaires (Schulze et al., 2021). But these effects can also be elicited by manipulating the instructions or items. The contextualization reduces within-person inconsistency and between-person variability of adopted frames of reference, and a congruent contextualization regarding the criterion has been found to have incremental validity compared to non- or incongruently contextualized ones (Bing et al., 2004; Lievens et al., 2008; Schmit et al., 1995; Shaffer & Postlethwaite, 2012).

The Present Study: Frame of Reference Effects in Faking Instructions

As mentioned above, we assume that the effects of induced faking in laboratory studies with applicant scenarios and noncontextualized questionnaires may be underestimated within both between- and within-subject designs because frame of reference effects have not been

controlled. We derived this assumption as follows: The generic instructions of personality questionnaires ask the participants to answer the questions spontaneously and with regard to their typical behavior. Within these instructions no context is provided and only the personal characteristics (Lievens et al., 2008) and the hidden frames of reference in certain items (Schulze et al., 2021) influence the mental search space. This leads to intraindividual variability and interindividual differences of adopted frames of reference (Lievens et al., 2008).

To elicit the motivation to fake, applicant scenarios are introduced and instruct the participants to answer the items as if they were applying for a job that they really want and to act in a way to maximize their chances of getting hired (e.g., Davison et al., 2021; Harris et al., 2020). We assume that these instructions may not only create a high stakes context – they may also trigger a work-related frame of reference. In this paradigm, faking and frame of reference effects would be confounded (see Figure 1).

As both faking and frame of reference effects have shown to be different in magnitude across individuals (Lievens et al., 2008; Schulze et al., 2021; McFarland & Ryan, 2000), the problem might be less pronounced in within-subject designs because these interindividual differences will be ruled out when comparing scores derived from the same person. But still, if the applicant instructions elicit both faking and frame of reference effects they will do so within an individual and both processes will be confounded intraindividually, making it necessary to test this assumption and to control for frame of reference effects if this assumption proves to be reasonable empirically.

In frame of reference research, it has been argued that there could be positive outcomes of contextualization regarding validity because of frame of reference effects but also negative outcomes because of faking. Bing et al. (2004) and Schmit et al. (1995) studied frame of reference effects and reported contextualized questionnaires to have higher and incremental predictive validity compared to noncontextualized ones. Thus, they concluded possible faking due to contextualization to be rather negligible (see also Bing et al., 2014; Robie et al., 2001). But they also found mean shifts between instruction types (generic vs. applicant) with applicant instructions producing more positive values (i.e., higher scores on extraversion, openness, conscientiousness, agreeableness, and lower scores on neuroticism) – findings similar to those reported by faking research (e.g., Birkeland et al., 2006).

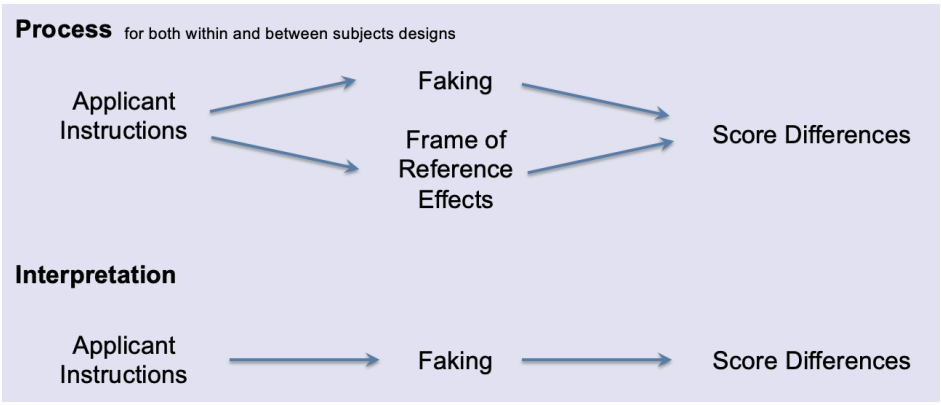

Figure 1. *Assumed processes and interpretation of laboratory faking studies for both between and within subject designs.*

The other way around, this assumption has, to our knowledge, not yet been tested. If the applicant instructions in faking research do already introduce a frame of reference, which has been shown to have a positive effect (e.g., Shaffer & Postlethwaite, 2012), and generic instructions do not, this may a) attenuate the effects of faking behavior and b) both effects would be confounded, and the criterion of internal validity as a requirement to draw conclusions regarding faking would not be met. In other words, our research question is: Do frame of

reference effects have to be controlled when studying faking good with applicant scenarios in laboratory settings to ensure internal validity?

Extensions as compared to previous research

Our design will resemble previous studies regarding sample and design (see the detailed description below, e.g., Schmit et al. 1995; Bing et al., 2004) to make our findings significant for the specific field of application. But we will adapt and extend in important ways, as follows: We will use a full-within repeated measurement design with a between-measurements interval to control for individual characteristics, possible transfer effects (Burns & Christiansen, 2006; Davison et al., 2021) and demand characteristics (Orne, 1962) which can produce systematically biased results. We will additionally control for demand characteristics by asking participants to make assumptions regarding the purpose of the study, which is an important extension and has, to our knowledge, not been assessed before. To reduce the number of individuals who find out about the real topic of the study we will work with a cover story and inform participants that we intend to evaluate instructions in the context of questionnaire construction. We discussed the possible objection of drawing attention to the instructions but argue that the instructions are intended to have an effect and are sometimes read rather carelessly (Marasi et al., 2019) – thus drawing attention to the instructions seems to be an advantage rather than a problem. Furthermore, we will ensure data quality by predefining strict exclusion criteria (see below).

As models of faking suspect experience to be an antecedent of faking behavior (Goffin & Boyd, 2009) and the adoption of a frame of reference should be easier or even solely possible when having experience in the relevant context (Schmit et al., 1995, Robie et al., 2000), we will use a clearly defined and well-known position in the applicant scenario for student participants – a student assistant job. Additionally, we will measure experience and knowledge concerning student assistant jobs. If participants do not know anything about these kinds of jobs the applicant scenario may not be appropriate (Robie et al., 2000) as they would have to guess, pretend, or take other situations into account by assuming how they might behave in a certain context.

We will also control for intelligence as faking models posit intelligence as an important factor for faking ability (e.g., Levashina & Campion, 2006) and meta-analytic findings indicate that intelligence and faking behavior, which will be one of our dependent variables, are positively related (Schilling et al., 2020). Also, there are meta-analytic findings regarding the collinearity of conscientiousness, another dependent variable in our study, and cognitive ability which emerges when participants are motivated to fake (Davison et al., 2021). Furthermore, cognitive ability is related with our third criterion variable average university grade and is therefore important to control for as has been done in previous research (Bing et al., 2004).

Following the rationale of frame of reference research and the theory of conditional dispositions (Wright & Mischel, 1987) it is very likely that self-monitoring will have an influence on trans-situational variability of behavior with individuals scoring high on self-monitoring adapt more to the situational requirements than those scoring low on self-monitoring. Studies using rather short and not very differentiated methods of measuring self-monitoring (e.g., Snyder & Gangesad, 1986) only partially found it to be positively related to the intention to fake (Klehe et al., 2012; Konradt et al., 2011; Lester et al., 2015). We will therefore include two scales which provide a more differentiated picture of different aspects of self-monitoring (Laux & Renner, 2002) to meet this limitation of previous research and find out more about possible influences on intraindividual variability.

We will use different dependent variables: First as another important extension to directly assess the frame of reference effect, we will collect information regarding the adopted

frames of reference as to this point as the dependent variable only assessed by Fisher et al. (2017) in a field sample and not yet in laboratory settings using applicant instructions. Fisher et al. (2017) assessed adopted frames of reference globally following the personality questionnaire with three items each asking for one possible frame of reference (social, school, work) including a fourth option for more than one frame of reference within an item. Following the rationale of hidden frames of reference (Schulze et al., 2021) and intermittent faking (Brown & Böckenholt, 2022) which imply that both faking and frame of reference effects are no uniform processes but vary from item to item, we decided to assess the adopted frame of reference on the item level. We argue that there are also advantages regarding measurement and comprehensibility as it seems to be rather difficult to estimate the amount of an adopted frame of reference globally across a number of items in retrospect as compared to report which situation was in one's mind when answering one specific item. Of course, it is possible that the repeated question will increase the attentiveness concerning the frame of reference but as we will always give different options including an empty response option, we argue that it is very unlikely to introduce systematic influence into a certain direction (see item wording below).

Second, we will use conscientiousness scores between conditions as a measure of changes in response behavior. We will focus on conscientiousness for different reasons: As previous research proceeded similarly (e.g., Bing et al., 2004; Robie et al., 2000), we are able to a) replicate previous findings and b) derive more specific hypotheses regarding changes in response behavior, which is crucial to conduct reliable statistical analyses (Wicherts et al., 2016). Furthermore, conscientiousness is specifically relevant for the application context as this dimension has proven to be predictive of job performance across different jobs (Barrick & Mount, 1991).

Frame of reference effects for conscientiousness have been found to be smaller compared to the effects on measures of the other big five dimensions, that is why it is assumed that even the noncontextualized items to assess conscientiousness elicit the adoption of an at work frame of reference rather than the noncontextualized items to assess openness, extraversion, neuroticism, or agreeableness (Schulze et al., 2021; Shaffer & Postlethwaite, 2012). In contrast to frame of reference effects, faking has been found to be highest for conscientiousness (Birkeland et al., 2006). Thus, if the frame of reference effect of applicant instructions does occur within conscientiousness measurement when frame of reference effects are low and faking effects are high, it is very likely that this finding will generalize to the remaining big five dimensions.

To collect information regarding the effect of contextualization to be expected for every item in advance, we will conduct a preliminary study similar to Robie et al. (2000) and Lievens et al. (2008) which will assess the situational association of each item in its contextualized and noncontextualized form. To address suspected limitations (Robie et al., 2000), we will collect information regarding the elicited frame of reference by both a student and an expert sample. We will also include a rating concerning item-relevance regarding a student assistance job to be able to make an educated assumption about the magnitude of item-specific faking effects. For this rating it will also be of advantage to rely on both a student and an expert sample to uncover possible perceptual differences regarding which behavior might be important for student assistants.

Following the rationale of work-relevance of conscientiousness (Barrick & Mount, 1991), we assume that the contextualization for conscientiousness items will be especially plausible for participants, which we think will increase acceptance and, together with the fact that the decision to focus on only one dimension leads to a decrease in questionnaire-length, will reduce dropout rates. At last, the reduction of items enables us to assess the adopted frame of reference and effects of response distortion not globally but on the item-level which

seems to be especially relevant following research regarding hidden frames of reference (Schulze et al., 2021) and intermittent faking (Brown & Böckenholt, 2022).

As our third dependent variable, we will collect information regarding average university grade as a criterion measure. Even though Lievens et al. (2008) found college GPA to be better predicted by an “at school”-contextualized questionnaire and not by an “at work”-contextualization, we decided to do so based on the following argumentation: Unlike Lievens et al. (2008) who manipulated the frame of reference without referring to a specific job to keep in mind, we introduce a student assistant job in our applicant scenario which is well known in the student population and by definition regarding content- and task-related requirements closely related to the university context. Thus, we argue that average university grade will serve as an appropriate criterion to assess and compare criterion-related validity of the different questionnaires.

Hypotheses

Changes in response behavior can be elicited by many different factors. Faking as the intentional response distortion (Röhner & Schütz, 2020) and frame of reference effects as adapted responses to account for situational changes in actual behavior (Shaffer & Postlethwaite, 2012) are two of these factors as discussed above. With the current study we first want to assess if applicant instructions directly influence the adopted frame of reference as compared to the generic instructions. If so, it proves the assumption to be right that frame of reference effects are inherent to applicant instructions and that these effects are therefore confounded with the motivation to fake also elicited by these instructions. We assume that the work-frame of reference will be reported with a higher frequency in the noncontextualized applicant instructions condition as compared to the noncontextualized generic instructions condition (H1).

We then want to replicate the finding that the item-level contextualization also influences the adopted frame of reference (Lievens et al., 2008, Fisher et al., 2017) to allow to introduce this manipulation to be a mean to influence frame of reference effects independently of faking and thus researchers may be able control for these effects. We therefore hypothesize that the work-frame of reference will be reported with a higher frequency in the contextualized conditions as compared to the noncontextualized conditions (H2).

As the instructions are only given once whereas the item-level contextualization is repeated for every single item constantly remembering participants to think of the work context, we expect the effect of item-level contextualization to be higher as compared to the effect of applicant instructions (see also Swift & Peterson, 2019) (H3). If this holds to be true, it can be counted as first evidence that the stronger item-level contextualization can adjust for the weaker instructional effect and lead to consistency of frame of reference effects between instruction conditions.

To evaluate the relevance of the frame of reference effects within applicant instructions, we will compare the magnitude of changes of the adopted frames of reference between the noncontextualized vs. contextualized generic instructions conditions with the noncontextualized vs. contextualized applicant instructions conditions. We expect the effect of item-level contextualization of the frequency of the work-frame of reference to be adopted to be significantly smaller between the applicant instructions conditions as compared to the generic instructions as the inherent frame of reference in the applicant instructions is already taking effect (H4).

To provide evidence that the item-level-contextualization might be a solution to separate faking elicited by applicant instructions from frame of reference effects, there should be no difference in the adopted frames of reference between instructional conditions ones the

questionnaire is contextualized. Thus, we assume that there will be no difference in the frequency the work-frame of reference is reported between the contextualized applicant instructions condition and the contextualized generic instructions condition (H5).

The assumptions derived above imply an order of conditions regarding the frequency the work frame of reference is reported. From lowest to highest we expect the following order: generic instructions without contextualization, applicant instructions without contextualization, generic instructions, and applicant instructions with contextualization (G- < A- < G+ = A+, for illustration see Figure 2) (H6).

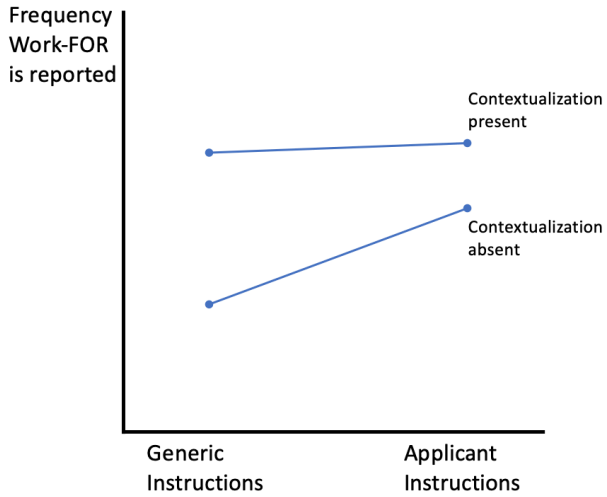

Figure 2. *Assumed effects of instructions and contextualization on the adoption of a work-frame of reference.*

After having assessed the direct effects of instructions and contextualization on the adopted frames of reference we want to examine the influence of both processes on the conscientiousness scores. We will do so because these scores are the critical variables used to predict relevant outcomes such as job performance and should not be biased to allow to derive reliable decisions e.g., within personnel selection. We first want to replicate the findings of former faking studies that applicant instructions lead to a change in mean scores as compared to generic instructions (Birkeland et al., 2006). We therefore hypothesize that within both the contextualized and noncontextualized conditions there will be a higher mean score of conscientiousness for the applicant instructions as compared to the generic instructions. Originally attributed to faking alone (e.g., Bensch et al., 2019), we argue that these changes are caused by both the instruction-inherent frame of reference effects and faking (H7).

To assess frame of reference effects independently of faking, we want to compare conscientiousness mean score changes between the contextualized and noncontextualized conditions for both the applicant and the generic instructions. Studies regarding frame of reference effects which found contextualized questionnaires to elicit a change in scale means reported similar results as faking studies (e.g., Schmit et al., 1995). But in contrast these changes are attributed to differences in behavioral samples drawn during a more selective memory search rather than a deliberate response distortion (Lievens et al., 2008). Following this rationale, it is plausible that if individuals think of the work context where conscious behaviors are valued and in addition to the individual tendency in general induced by the situation rather than in the private context, they will self-report more conscientiousness (see also Shaffer & Postethwaite, 2012). Thus, we assume that contextualized conditions produce a higher conscientiousness mean score as compared to noncontextualized conditions within both applicant and generic instructions conditions (H8).

As derived above, frame of reference effects for the dimension of conscientiousness have been found to be smaller (Schulze et al., 2021; Shaffer & Postlethwaite, 2012) and faking as found to be higher (Birkeland et al., 2006) as compared to the remaining big five conditions. We therefore expect the effect of instructions on conscientious mean scores to be higher than the effects of item-level-contextualization (H9).

If both the faking and frame of reference effects lead to change in conscientiousness mean scores and both processes are confounded within the applicant instruction, some of the changes between the noncontextualized generic instructions condition and the noncontextualized applicant instructions condition should be attributed to frame of reference effects rather than faking. Or to put it the other way around, if both frame of reference and faking effects are elicited by applicant instructions and both effects lead to an increase in mean scores, the effect of the instructions should be greater in the noncontextualized than in the contextualized conditions, as in the contextualized the frame of reference effects are already accounted for by the item-level contextualization. Consequently, we predict the conscientiousness mean score changes between the noncontextualized generic instructions vs. applicant instructions conditions to be greater than changes between the contextualized generic instructions vs. applicant instructions conditions (H10).

The assumptions derived above imply an order of conditions regarding change in conscientious mean scores. From lowest to highest we expect the following order: generic instructions without contextualization, generic instructions with contextualization, applicant instructions without contextualization, and applicant instructions with contextualization (G- < G+ < A- < A+, for illustration see Figure 3) (H11).

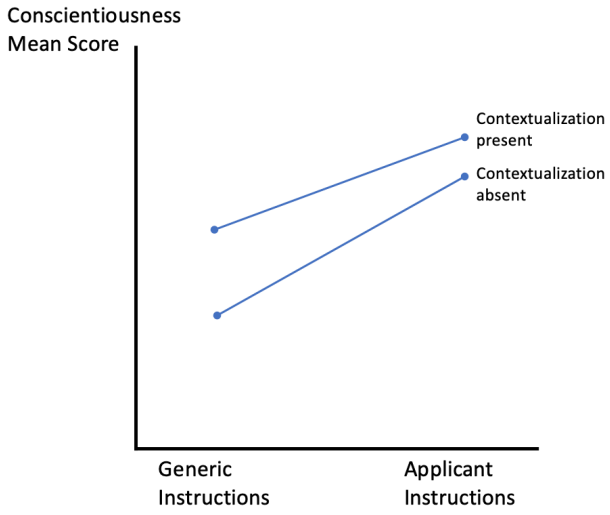

Figure 3. *Assumed effects of instructions and contextualization on the conscientiousness mean scores.*

Finally, we want to examine the implications regarding criterion-related validity. The frame of reference research has found criterion-congruently contextualized questionnaires to yield a higher criterion related validity than noncontextualized or criterion-incongruently contextualized ones as interindividual differences and intrainividual variability is reduced and conceptual overlap between items and criterion is increased (Lievens et al., 2008). At the same time faking research produced somewhat mixed findings (e.g., Marcus, 2009, Ziegler et al., 2012) but there are arguments leading to the assumption that e.g., rank order changes (Birkeland et al., 2006) due to deliberately distorted answers might be a danger for subsequent decision making (e.g., Schmitt & Bradburn, 2018). To simplify, frame of reference effects and their consequences are rather desirable whereas faking and its potential consequences are

rather undesirable. If both processes are confounded within the applicant instructions it might lead to the underestimation of faking effects on the criterion related validity, i.e., the prediction of relevant outcomes.

Following this rationale, we expect contextualized questionnaires to lead to a higher criterion related validity as compared to noncontextualized questionnaires (H12). If faking does have a negative effect, and frame of reference effects conversely do have a positive influence regarding validity, we assume only a small albeit negative (Schmit et al., 1995, Bing et al., 2004) effect on the criterion related validity of the applicant instructions as compared to the generic instructions within the noncontextualized conditions (H13). As contextualized questionnaires have repeatedly been shown to increase validity even in the application context (e.g., Lievens et al., 2008), we assume that the frame of reference effects will be higher that the faking effects (H14). If frame of reference effects are held constant, the faking effect alone will be the cause of changes in criterion related validity. We therefore hypothesize that within the contextualized conditions the applicant instructions as compared to the generic instructions will have a more pronounced negative effect than within the noncontextualized condition (H15). Following these assumptions, we expect the following order of conditions regarding criterion-related validity from lowest to highest: applicant instructions without contextualization (as there are faking effects and intraindividual variability of frames of reference), generic instructions without contextualization (no faking but variability in frames of reference), applicant instructions with contextualization (faking effects but increased within-person consistency of adopted frames of reference), and generic instructions with contextualization (no faking and increased within-person consistency of adopted frames of reference) (A- < G- < A+ < G+, for illustration see Figure 4) (H16).

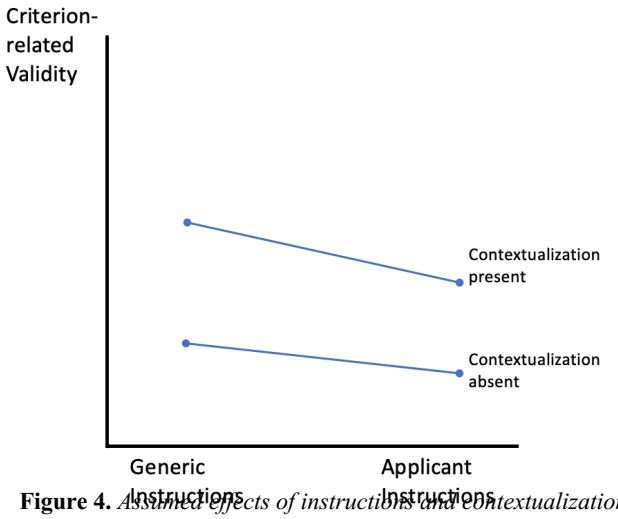

Figure 4. *Assumed effects of instructions and contextualization on criterion-related validity.*

Methods

Transparency, Openness and Reproducibility

Theory, hypotheses, and methods will be preregistered prior to the analyses. Furthermore, R code will be made available after the analyses.

Design

To answer the research question and test the hypotheses, we plan to conduct a 4-wave longitudinal study with a 2x2 full-within repeated measurement design, as within-subject designs allow to control for individual characteristics in response behavior and have been

found to produce more accurate estimates (Viswesvaran & Ones, 1999). There will be an interval of approximately 2 weeks between measurements which seems to be a reasonable compromise to reduce both transfer effects and dropout rates (Burns & Christiansen, 2006; Davison et al., 2021). There will be four conditions (see Table 1) as we will vary the instructions (generic ("honest") vs. application ("faking")) and the contextualization (item-level contextualization absent vs. present).

Table 1

Conditions and assumed causes for conscientiousness mean differences between conditions.

	Item-level contextualization present (+)	Item-level contextualization absent (-)	Causes for differences between conditions
Generic instructions, "honest" (G)	G+	G-	Frame of reference effects in generic instructions
Applicant instructions "faking" (A)	A+	A-	Frame of reference effects in applicant instructions
Causes for differences between conditions	Faking	Faking + frame of reference effects	

Note: *G- (generic instructions, contextualization absent), G+ (generic instructions, contextualization present), A- (applicant instructions, contextualization absent), A+ (applicant instructions, contextualization absent).*

To be able to control for and vary the frame of reference independent of faking behavior, we will use item level contextualization as a manipulation, an approach that has been used successfully in past research (e.g., Bing et al., 2004). We will assess mean differences between conditions and include a measurement of the adopted frame of reference as a dependent variable. Moreover, we will include self-monitoring, general mental abilities, and experience and knowledge regarding the job from the applicant instructions.

To ensure good data quality, we will incorporate attention items, measurements regarding motivation and present context while answering the questionnaire and collect participant assumptions regarding the purpose of the study.

Sample

We will limit participants to the German speaking student population. We will survey students from all disciplines to increase variance, an approach that has been used in previous research (e.g., Davison et al., 2021, Bing et al., 2004). Thus, our results can be compared with the existing literature.

As we intend to assess criterion related validity by means of correlations, we used G*Power to calculate the needed sample size to infer a cautiously expected correlation of .25 (meta-analytic correlation of conscientiousness and GPA, see Vedel, 2014: .26; see Mammadov, 2022: .27), given the conventional alpha of .05 and beta of .8. G*Power yielded an appropriate sample size of *N*=95.

We repeated this procedure for a repeated measures ANOVA, with within-subject repetition at 4 measurement waves. To define an expected effect size, we consulted the literature. Birkeland et al. (2006) meta-analyzed studies comparing noncontextualized "honest" and "faking" conditions and reported the highest effect size for mean differences (d=.45, equivalent to eta^2 = .048 according to Lenhard & Lenhard, 2016) for conscientiousness. For contextualized vs. noncontextualized questionnaires no meta-analytic effect sizes were available for mean differences. We therefore extracted effect sizes from different studies (Robie et al., 2000: d =.52, equivalent to eta^2 = .063 according to Lenhard & Lenhard, 2016;

Lievens et al., 2008: eta^2 = .03; Schmit et al., 1995: eta^2 = .15, .11). All these effect sizes considered, we calculated the needed sample size for an cautiously expected eta^2 of .05. G*Power yielded an intended sample size of *N*=28, given the described effect size, the conventional alpha of .05 and power of .8, and a correlation of .5 between measurements, conservatively expected after consulting the literature (e.g., Heller et al., 2009; Wang et al., 2021).

These sample sizes differ considerably from each other. But we want to assess the criterion related validity of the questionnaires of the different conditions with enough power, being able to detect significant changes. Furthermore, we not only want to detect mean changes between conditions but also differentiate these changes into faking and frame of reference effects. To put it clearly, we plan not only to find but also to subdivide the expected effects. That is why we intend to oversample regarding the mere detection of mean changes and decided to aim at the sample size of N=95.

To be certainly included in the final analysis participants will have to take part in all four waves of measurement. They will have to answer all attention items [1](e.g., "*Please answer this item with completely agree."*), correctly. We will also include 4-5 items asking participants if they completed the questionnaire carefully, took their time, participated in a quiet environment, or if they hurried (inverted). If the applicant instructions were given, an additional item asked if individuals could emphasize with the situation. They will have to rate each item with at least "*rather agree"* (for the inverted item "*rather not agree*", on a five-point scale, see all options in the appendix).

Table 2

Coding system of Blindness regarding the purpose of the study

0: blind *(no markers mentioned)*
1: non-blind *(one or more markers mentioned)*
• Answering consistently • Behavioral difference between contexts • Chance across time/ contexts • Stability of personality • Ability to adapt oneself (*ability to emphasize with the situation* is not a marker) • Faking • Social desirability/SDR • Lying/ detection of lies • Impression management • Distortion of answers • Biased self-presentation

To control potential demand characteristics, we not only chose the interval of two weeks between waves (see above), but we will also ask participants to make assumptions regarding the objectives of the study. If they mention predefined markers (see Table 2) they will be coded as non-blind to the study's purpose. For each quality criterion we will analyze subgroup differences. If these analyses reveal group differences, participants not meeting the criteria will be excluded from the final analyses.

Based on previous research (e.g., Davison et al., 2021; Lee et al., 2019; Liu & Zhang, 2020; Roess & Roche, 2017, Terry et al., 1999), we expect a dropout rate of about 20%.

[1] We chose to use one attention item for every 20 items within a page of the questionnaire as attention may vary over the course of the study and thus one item would not be a sufficient indicator (Edwards, 2019).

Moreover, we expect that about 10% of the participants will meet the exclusion criteria. Thus, we will further oversample and increase the number of necessary participants for t1, aiming to collect N=300 datasets.

Measures

The following variables will be recorded during the four waves of the study. See the measurements and item wording of all variables in German in the questionnaires in the appendix.

Independent Variables. The independent variables will be a) one of two instructions (generic[2]: "*Below are several characteristics that may apply to you. For each of the following statements below, please indicate the extent to which you agree. There are no right or wrong answers. Always choose the answer that describes you best. Please be as honest as possible."* vs. applicant: *"You are applying for a very popular and well-paid student assistant job with long-term perspective. Below you will find several characteristics that could apply to you in this job. For each of the following statements below, please indicate the extent to which you agree. Answer the following questionnaire in such a way that you show your best side to increase your chances of getting the job."*) and b) the presence or absence of the item-level contextualization implemented by adding an "*at work*" tag to the items of the BFI-2 (Danner et al., 2016) always following the verb of the item (see different conditions in Table 1 and procedure and order of conditions within groups in Table 3).

Dependent Variables. The dependent variables will be the changes in the conscientiousness scores, the adopted frames of reference and the current *average university grade* which will be collected together with the demographic variables.

The *adopted frames of reference* will be measured with four items translated and adapted from Fisher et al. (2017). To assess the adopted frame of reference on the item level we will include these items after each of the 12 conscientiousness items. Participants have to indicate (yes/ no) if they thought of a specific context (social, school, work) while answering the previous item. We will include a fourth option with an empty field where participants will be able to add a missing context, to ensure that participant will not recognize that we are only interested in the adoption of the work frame of reference.

Conscientiousness will be measured with 12 Items from the BFI-2 (Danner et al., 2016). Reliability in previous research was good with Cronbach's alpha of .86 and acceptable to good for the facets with Cronbach's alpha of .84 for Organization, .72 for Productiveness and .58 for Responsibility.

Control Variables. In addition to demographic information (age, gender, university, course of study, average grade), we will collect information regarding the following control variables which have been shown to influence faking behavior and the amount of contextual variability of individual behavior. First, we will assess *self-monitoring and concern for appropriateness* using the Revised Self-Monitoring Scale (reported Alpha = .8, 12 items) and the Concern for Appropriateness Scale (reported Alpha = .85, 12 items) (Laux & Renner, 2002). Also, we will control for general mental abilities measured with the 20-item version of the Hagen Matrices Test (Heydasch, 2014). To assess *experience and knowledge*, we created 10 items asking for own experiences, experiences conveyed by third parties, theoretical knowledge, and the attitude regarding student assistant jobs.

[2] We adapted the original instruction by excluding the example ("For example, would you say about yourself that you enjoy spending time with other people?") as this may trigger a private frame of reference (see also Shaffer and Postlewaite, 2012). We also added the last two sentences, which ask the participants to answer honestly to meet the common "honest condition" used in faking research (Shaffer & Postlewaite, 2012).

Procedure

After the baseline measurement at wave 1 participants will be randomly assigned to one of six groups. The groups will differ in the order in which the different questionnaire conditions of the BFI-2 conscientiousness scale (Danner et al., 2016) (based on the 2x2 design: instructions: generic vs. applicant; item-level contextualization: preset vs. absent, see Table 1) will be administered (see Table 3). This is done to rule out sequence effects as it has been shown that the order of conditions influences the magnitude of faking effects (Hu & Connelly, 2021).

Table 3

Procedure: Order of conditions within groups and control variables measured at each wave of measurement

	Wave 1	Wave 2	Wave 3	Wave 4
Group 1	G-	G+	A+	
Group 2			A-	
Group 3		A+	G+	
Group 4			A-	
Group 5		A-	G+	
Group 6			A+	
Control variables assessed at wave i	- Demographics - Self-Monitoring - GMA - Adopted FOR - Motivation & Environment	- Adopted FOR - Motivation & Environment	- Adopted FOR - Motivation & Environment	- Adopted FOR - Motivation & Environment - Experience & Knowledge - Assumptions regarding purpose of the study

Note: *G- (generic instructions, contextualization absent), G+ (generic instructions, contextualization present), A- (applicant instructions, contextualization absent), A+ (applicant instructions, contextualization absent).*

The first wave will be the baseline measurement, the same for all participants and the longest questionnaire (approximately 20-30 minutes). Participants will first be given information regarding the study and will receive a privacy statement. Only if they provide informed consent, they will be transferred to the questionnaire. Individuals will then generate a code, which will be used to link the questionnaires of each person while keeping their anonymity. Following this, we will collect demographic information (gender, age, course of study, university, average university grade), the baseline conscientiousness measurement (G-), the adopted frames of reference, self-monitoring, general mental abilities, and some control items regarding motivation and environment during participation. At the end of the questionnaire, we will collect an e-mail-address to contact participants for the upcoming waves.

At wave 2 to 4 participants will receive questionnaires each taking 10 to 15 minutes. They will be given the remaining conditions of the BFI-2 conscientiousness scale (Danner et al., 2016) in the order given by their assigned group (e.g., an individual assigned to group 4 will receive questionnaire A+ at wave 2, A- at wave 3, and G+ at wave 4).

At wave 2 and 3 only the conscientiousness and the adopted frames of reference will be measured. At wave 4 we will additionally collect information regarding the experience/ knowledge and assumptions concerning the purpose of the study. Finally, participants will be given the choice between participation in a lottery to win one of six vouchers (1x50, 4x20 und 5x10 Euro for a sustainable online marketplace or for local stores in Bamberg) or individual feedback regarding their personality profile (based on the baseline measurement). Students from the University of Bamberg can additionally receive course credit.

Confirmatory and Exploratory Analyses Plan

Preparation and Subgroup-Analyses

All our analyses will be conducted using the software R (R Core Team, 2019). We will begin our analyses by preparing our data. First, we will download our data from soscisurvey (https://www.soscisurvey.de/), create a new dataset including only the relevant variables, bring together individual data using the participant's self-generated code, and create a new variable for group and condition. Then, after excluding every participant who did not provide full information in the first questionnaire, we will aggregate data from the baseline measurement (wave 1), i.e., calculating scale means or sum scores, reliabilities, and measures of dispersion for baseline conscientiousness, adopted frames of reference, self-monitoring and concern for appropriateness, cognitive ability, knowledge and experience, motivation and environment and attention items.

Next, we will analyze missing data in the subsequent waves of measurement, i.e., assessing if there are any missing data, examining if there are obvious missing data patterns – as we conduct a longitudinal study, we expect to see a monotone pattern with an increase in participant loss throughout the study – and trying to predict missingness by our baseline variables. If the probability to drop out of the study correlates e.g., with baseline conscientiousness, this will limit the generalizability of our analyses. Every participant with missing data will be excluded.

After adjusting for drop out we will aggregate the data of waves 2-4 analogous to wave 1, additionally coding the assumptions regarding the purpose of the study as blind or non-blind. Then we will calculate difference variables for the repeated measures.

We will use the difference variables to conduct subgroup analyses to determine if blindness, knowledge and experience or the score for the attention items influences scale means or group differences. If we find group differences, we will exclude participants not meeting the relevant quality criteria.

Descriptive Statistics and Test of Requirements

After selecting the final sample, we will calculate descriptive statistics regarding demographic information, dependent, independent and control variables (means, reliabilities, standard deviations) and visualize sample distributions and correlations with tables and box-plot diagrams.

The next step in our analyses is to determine if the requirements for subsequent statistical calculations are met. As the main method will be the repeated measures ANOVA these requirements include the absence of outliers (as can be seen in the box plot diagrams) independent, identically normally distributed residuals, individual effects, and interaction effects, and derived from these assumptions: sphericity (which can and will be tested with the Mauchly-test).

Hypotheses Testing

To give an overview, we will use 2x2 repeated measurement ANOVAs (Model 1) to examine both the effects of instructions and contextualization on the adopted frames of reference and conscientiousness scores and criterion related validity. Generalized ANCOVAs (Model 2) will be used to include the above-mentioned control and demographic variables and assess the robustness of the effects. Dependent sample t-tests will be used to compare the magnitude of effect sizes. Contrast analyses will be used to test the trend hypotheses.

The equations for the ANOVAs and ANCOVAs will be defined with i being the marker for respondents (respondent 1, 2, 3, …, n), y_i being the dependent variable score (adopted frames

of reference, conscientiousness, or criterion related validity), x_i representing a vector of a variable influencing the dependent variable scores and β_i the associated regression coefficient. See the definitions of each variable and their regression coefficient in Table 4. ε_i will be the error term.

Model 1

$$y_i = x_{i1}\beta_{i1} + x_{i2}\beta_{i2} + x_{i1}x_{i2}\beta_{i3} + \varepsilon_i$$

Model 2

$$y_i = x_{i1}\beta_{i1} + x_{i2}\beta_{i2} + x_{i1}x_{i2}\beta_{i3} + x_{i4}\beta_{i4} + x_{i5}\beta_{i5} + x_{i6}\beta_{i6} + x_{i7}\beta_{i7} + x_{i8}\beta_{i8} + \varepsilon_i$$

Table 4

AN(C)OVA vectors and their associated variables.

Regression vector	Variable
$x_{i1}\beta_{i1}$	Factor 1: instructions (generic vs. faking)
$x_{i2}\beta_{i2}$	Factor 2: contextualization (present vs. absent)
$x_{i1}x_{i2}\beta_{i3}$	Interaction-term for factorial variables
$x_{i4}\beta_{i4}$	Self-monitoring (self-presentation competence, perception sensitivity)
$x_{i5}\beta_{i5}$	Concern for appropriateness (protective variability, protective social comparison)
$x_{i6}\beta_{i6}$	General mental abilities
$x_{i7}\beta_{i7}$	Age
$x_{i8}\beta_{i8}$	Gender

Adopted Frame of Reference (Hypotheses 1 – 6)

To assess effects of contextualization and instructions regarding the adopted frame of reference, we will first apply a 2x2 repeated measurement AN(C)OVA with the adopted frame of reference as the dependent variable and instruction type and contextualization as factors with two levels each. Put statistically, we expect significant main effects for both instructions (H1) and item-level contextualization (H2) and a significant interaction effect as first evidence for hypotheses 3 to 6. Hypothesis 3 to 5 will be tested using t-tests for dependent samples to compare a) mean effect sizes of contextualization and instructions (H3), b) the effects of contextualization within the generic instructions condition and the effect of contextualization within the applicant constructions conditions (H4), and c) the difference of the instructions conditions if item-level contextualization is present (H5). Finally, a contrast analysis will be conducted to test the expected order of conditions as defined in Hypothesis 6 (see Table 5).

Table 5

Contrasts predefined to test the trend hypothesis (H6)

	X_1	X_2	X_3
G-	$\frac{1}{2}$	$\frac{1}{2}$	0
A-	$\frac{1}{2}$	$-\frac{1}{2}$	0
G+	$-\frac{1}{2}$	0	$\frac{1}{2}$
A+	$-\frac{1}{2}$	0	$-\frac{1}{2}$

Note: *G- (generic instructions, contextualization absent), G+ (generic instructions, contextualization present), A- (applicant instructions, contextualization absent), A+ (applicant instructions, contextualization absent).*

Conscientious Mean Scores (Hypotheses 7-11)

Analogous to the hypotheses regarding the adopted frame of reference, we will conduct another AN(C)OVA to test the main effects of instructions (H7), contextualization (H8) and interaction effects (H 9-11) on the conscientiousness mean scores as the dependent variable. T-tests will be applied to test the assumptions regarding effect sizes of the main effects (H9) and the effect of instructions within the noncontextualized as compared to the contextualized conditions (H10). The trend hypothesis (H11) will be tested using a contrast analysis (see Table 6).

Table 6

Contrasts predefined to test the trend hypothesis (H11)

	X_1	X_2	X_3
G-	$\frac{3}{4}$	0	0
G+	$-\frac{1}{4}$	$\frac{2}{3}$	0
A-	$-\frac{1}{4}$	$-\frac{1}{3}$	$\frac{1}{2}$
A+	$-\frac{1}{4}$	$-\frac{1}{3}$	$-\frac{1}{2}$

Note: *G- (generic instructions, contextualization absent), G+ (generic instructions, contextualization present), A- (applicant instructions, contextualization absent), A+ (applicant instructions, contextualization absent).*

Criterion-Related Validity (Hypotheses 12-16)

Finally, we will assess the criterion related validity by the examination of correlations of the conscientiousness scores with the average university grade as the dependent variable, again by conducting an AN(C)OVA first. We expect to obtain significant main effects for contextualization (H12) and instructions (H13) and interaction effects (H14-16). We will conduct t-tests regarding main effects of contextualization as compared to instructions (H14) and the effects of instructions within the contextualized as compared to the effect of instructions within the noncontextualized condition (H15). And again, we will test the trend hypothesis (H16) by applying a contrast analysis (see Table 7).

Table 7

Contrasts predefined to test the trend hypothesis (H16)

	X_1	X_2	X_3
A-	$\frac{3}{4}$	0	0
G-	$-\frac{1}{4}$	$\frac{2}{3}$	0
A+	$-\frac{1}{4}$	$-\frac{1}{3}$	$\frac{1}{2}$
G+	$-\frac{1}{4}$	$-\frac{1}{3}$	$-\frac{1}{2}$

Note: *G- (generic instructions, contextualization absent), G+ (generic instructions, contextualization present), A- (applicant instructions, contextualization absent), A+ (applicant instructions, contextualization absent).*

Exploratory Analyses

An important extension of the current study is the measurement of the adopted frame of reference on the item-level. To our knowledge this has not been done before. That is why we have several assumptions and questions we want to include as exploratory analyses as there is not enough literature yet to derive profound hypotheses including effect size estimates to conduct a priory power analyses and sample planning.

First, it would be interesting to assess if individuals as well as items differ in terms of their tendency to adopt or elicit a certain frame of reference. This is assumed by most studies investigating frame of reference effects (e.g., Robie et al., 2000; Schulze et al., 2021) but has not been addressed by research to date. We will analyze the data of our preliminary study as well as the baseline measurement of the main study to address this question. Also, it may be interesting to explore, which variables may influence this individual tendency to adopt a certain frame of reference. That is, why we want to search our data for interesting patterns regarding the collected variables.

As the frame of reference induced by item-level contextualization should limit the mental search space to work related memories of one´s behavior (e.g., Lievens et al., 2008) and therefore establish a common basis for all scores one can assume that frame of reference effects will lead to a) an increase in internal consistency of the scale (see also Lievens et al., 2008) and b) rather uniform effects on the item-level, of course exhibiting ceiling effects for particularly high conscientiousness scores. Of course, these effects can only emerge if there is an actual change in the adopted frame of reference, which can be controlled as we measured the adopted frame of reference on the item level.

Another interesting question which could not be ruled entirely out to date is the emergence of faking due to contextualization. If contextualization of items only influences the adopted frame of reference, there should be no change in scores for those items which already elicited a work-frame of reference in the baseline measurement. If the contextualization does indeed motivate participants to fake, there should be score shifts for those items too – an assumption we can and intend to address with the given dataset.

Potential Implications

If the above elaborated assumptions prove themselves empirically, this can have important implications for past and future laboratory research regarding faking behavior. Past research may have to undergo reevaluation and replication, as the derived conclusions regarding faking may be biased due to a lack of internal validity. For future research this may mean, that frame of reference effects have to be ruled out or controlled for. There may be different options to do so which will have to be tested empirically:

First, if contextualization proves to unify the adopted frames of reference across instruction conditions, the use of contextualized questionnaires may be an option. In this study we chose to ad a context tag, which has been successfully used before (e.g., Robie et al., 2000) Nevertheless, there is also research regarding different and maybe even more precise ways to contextualize a questionnaire initially developed for general use (e.g., Wang et al., 2021). Of course, there are also questionnaires developed to assess personality in a certain context in the first place (e.g., Gill & Hodgkinson, 2007), but contextualization by means of a simple manipulation such as adding a tag to the item may be a more efficient and versatile way to improve validity (Shaffer & Postlethwaite, 2012). But it has to be noted that there is sometimes difficulty regarding adding a tag if it does not fit the subject initial items cover (Schulze et al., 2021), thus there are also approaches to completely contextualize questionnaires which yielded promising results but was again a very time- and resources-consuming manipulation (Holtrop et al., 2014).

Second, the implementation of preliminary studies may prove to be useful. Using an independent sample, items could be assessed regarding their hidden frame of reference within different instruction conditions – but, if individuals differ in their tendency to adopt a certain frame of reference, this may be a source of inaccuracy if there are sample differences. Nevertheless, we wanted to mention this option, as it may be a possibility to reassess past research by rating the used items regarding their focal frame of reference depending on instructions and explore if there are systematic differences due to changes in this item-characteristic.

Third, the usage of new “honest” instructions, which already imply the application context may be considered. For example, participants could be instructed as follows: *“You apply for a student assistant job (or any other) and are asked to fill out a personality questionnaire as part of the recruitment process. As you know that individuals are much more satisfied with a job that fits their personality and you do want to be satisfied with your*

occupation, you want to respond this questionnaire as honestly as you can, so that you are only selected if you fit the job and it fits you."

An according "faking" instruction might be: *"You apply for a student assistant job (or any other) and are asked to fill out a personality questionnaire as part of the recruitment process. You really want the job as you urgently need an occupation. As you know that only those individuals are selected whose personality fits the job profile, you want to fill out the questionnaire in a kind of manner that increases your chances of being selected."*

Of course, instructions like these will have to prove their usability within the laboratory context to ensure that internal validity is established. This as well as the examination of exploratory findings may be a task to be assigned to possible future research, given that this study yields the expected results.

References

Anglim, J., Molloy, K., Dunlop, P. D., Albrecht, S. L., Lievens, F., & Marty, A. (2021). Values assessment for personnel selection: comparing job applicants to non-applicants. *European Journal of Work and Organizational Psychology*. https://doi.org/10.1080/1359432X.2021.2008911

Barrick, M., & Mount, M. K. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Bensch, D., Maaß, U., Greiff, S., Horstmann, K. T., & Ziegler, M. (2019). The Nature of Faking: A Homogeneous and Predictable Construct? *Psychological Assessment*, *31*(4), 532–544. https://doi.org/10.1037/pas0000619

Bill, B., Melchers, K. G., Buehl, A. K., & Wank, S. (2020). An Investigation of Situational and Dispositional Antecedents of Faking Intentions in Selection Interviews. *Frontiers in Psychology*, *11*, 2034. https://doi.org/10.3389/fpsyg.2020.02034

Bing, M. N., Davison, H. K., & Smothers, J. (2014). Item-level Frame-of-reference Effects in Personality Testing: An investigation of incremental validity in an organizational setting. *International Journal of Selection and Assessment*, *22*(2), 165–178. https://doi.org/10.1111/ijsa.12066

Bing, M. N., Davison, H. K., Whanger, J. C., & VanHook, J. B. (2004). Incremental Validity of the Frame-of-Reference Effect in Personality Scale Scores: A Replication and Extension. *Journal of Applied Psychology*, *89*(1), 150–157. https://doi.org/10.1037/0021-9010.89.1.150

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A Meta-Analytic Investigation of Job Applicant Faking on Personality Measures. *International Journal of Selection and Assessment*, *14*(4), 317–335. Retrieved from http://www.researchgate.net/publication/227626532

Brown, A., & Böckenholt, U. (2022). Intermittent Faking of Personality Profiles in High-Stakes Assessments: A Grade of Membership Analysis. *Psychological Methods*. https://doi.org/10.1037/met0000295

Bourdage, J. S., Schmidt, J., Wiltshire, J., Nguyen, B., & Lee, C. (2020). Personality, interview faking, and the mediating role of attitudes, norms, and perceived behavioral control. *International Journal of Selection and Assessment*, *28*(2), 163–175. https://doi.org/10.1111/ijsa.12278

Burns, G. N., & Christiansen, N. D. (2006). Sensitive or Senseless: On the Use of Social Desirability Measures in Selection and Assessment. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 113–148). https://doi.org/10.1034/j.1600-0706.2001.950116.x

Calanna, P., Lauriola, M., Saggino, A., Tommasi, M., & Furlan, S. (2020). Using a supervised machine learning algorithm for detecting faking good in a personality self-report. *International Journal of Selection and Assessment*, *28*(2), 176–185. https://doi.org/10.1111/ijsa.12279

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., John, O. P. (2016). Die deutsche Version des Big Five Inventory (BFI).

Zusammenstellung Sozialwissenschaftlicher Items Und Skalen, *2*(2016), 121–132. https://doi.org/10.6102/zis247

Davison, H. K., Kluemper, D. H., Tao, S., Stewart, D. W., & Bing, M. (2021). The effects of faking on the relationship between cognitive ability and conscientiousness: A cautionary note. *International Journal of Selection and Assessment*, (February), 1–16. https://doi.org/10.1111/ijsa.12319

De Page, L., & Merckelbach, H. (2021). Associations between supernormality ('faking good'), narcissism and depression: An exploratory study in a clinical sample. *Clinical Psychology and Psychotherapy*, *28*, 182–188. https://doi.org/10.1002/cpp.2500

Detrick, P., & Roberts, R. M. (2021). Police applicant response bias on the California Psychological Inventory. *Psychological Services*. https://doi.org/10.1037/ser0000524

Edwards, J. R. (2019). Response invalidity in empirical research: Causes, detection, and remedies. *Journal of Operations Management*, *65*(1), 62–76. https://doi.org/10.1016/j.jom.2018.12.002

Ellingson, J. E., & McFarland, L. A. (2011). Understanding faking behavior through the lens of motivation: An application of vie theory. *Human Performance*, *24*(4), 322–337. https://doi.org/10.1080/08959285.2011.597477

Fisher, D. M., Cunningham, S., Kerr, A. J., & Allscheid, S. P. (2017). Contextualized personality measures in employee selection: Extending frame-of-reference research with job applicant samples. *International Journal of Selection and Assessment*, *25*(1), 18–35. https://doi.org/10.1111/ijsa.12156

Geiger, M., Bärwaldt, R., & Wilhelm, O. (2021). The Good, the Bad , and the Clever : Faking Ability as a Socio-Emotional Ability ? *Journal of Intelligence*, *9*(13).

Geiger, M., Olderbak, S., Sauter, R., & Wilhelm, O. (2018). The "g" in faking: Doublethink the validity of personality self-report measures for applicant selection. *Frontiers in Psychology*, *9*, 1–15. https://doi.org/10.3389/fpsyg.2018.02153

Gill, C. M., & Hodgkinson, G. P. (2007). Development and validation of the Five-factor Model Questionnaire (FFMQ): An adjectival-based personality inventory for use in occupational settings. *Personnel Psychology*, *60*(3), 731–766. https://doi.org/10.1111/j.1744-6570.2007.00090.x

Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology*, *50*(3), 151–160. https://doi.org/10.1037/a0015946

Goffin, R. D., & Woods, D. M. (1995). Using Personality Testing for Personnel Selection: Faking and Test-Taking Inductions. *International Journal of Selection and Assessment*, *3*(4), 227–236. https://doi.org/10.1111/j.1468-2389.1995.tb00035.x

Griffith, R. L., Yoshita, Y., & Chmielowski, T. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, *36*(3), 341–355. https://doi.org/10.1108/00483480710731310

Harris, A. M., McMillan, J. T., & Carter, N. T. (2020). Test-taker reactions to ideal point measures of personality. *Journal of Business and Psychology*, (2004). https://doi.org/10.1007/s10869-020-09682-8

Hartman, N. S., & Lee Grubb, W. (2011). Deliberate faking on personality and emotional intelligence measures. *Psychological Reports*, *108*(1), 120–138. https://doi.org/10.2466/03.09.28.PR0.108.1.120-138

Heggestad, E. D. (2012). A Conceptual Representation of Faking. Putting the Horse Back in Front of the Cart Eric. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New Perspectives on Faking in Personality Assessment* (pp. 87–101). https://doi.org/10.1093/acprof

Heller, D., Ferris, D. L., Brown, D., & Watson, D. (2009). The influence of work personality on job satisfaction: Incremental validity and mediation effects. Journal of Personality, 77(4), 1051–1084. https://doi.org/10.1111/j.1467-6494.2009.00574.x

Heydasch, T. (2014). *The Hagen Matrices Test (HMT) (part of doctoral dissertation thesis).* University of Hagen. https://doi.org/10.13140/RG.2.2.31433.75361

Holtrop, D., Born, M. P., de Vries, A., & de Vries, R. E. (2014). A matter of context: A comparison of two types of contextualized personality measures. *Personality and Individual Differences*, *68*, 234–240. https://doi.org/10.1016/j.paid.2014.04.029

Hu, J., & Connelly, B. S. (2021). Faking by actual applicants on personality tests: A meta-analysis of within-subjects studies. *International Journal of Selection and Assessment*, *29*(3–4), 412–426. https://doi.org/10.1111/ijsa.12338

Jeong, Y. R., Christiansen, N. D., Robie, C., Kung, M. C., & Kinney, T. B. (2017). Comparing applicants and incumbents: Effects of response distortion on mean scores and validity of personality measures. *International Journal of Selection and Assessment*, *25*(3), 311–315. https://doi.org/10.1111/ijsa.12182

Klehe, U. C., Kleinmann, M., Hartstein, T., Melchers, K. G., König, C. J., Heslin, P. A., & Lievens, F. (2012). Responding to Personality Tests in a Selection Context: The Role of the Ability to Identify Criteria and the Ideal-Employee Factor. *Human Performance*, *25*(4), 273–302. https://doi.org/10.1080/08959285.2012.703733

König, C. J., Merz, A. S., & Trauffer, N. (2012). What is in Applicants' Minds When They Fill Out a Personality Test? Insights from a qualitative study. *International Journal of Selection and Assessment*, *20*(4), 442–452. https://doi.org/10.1111/ijsa.12007

Konradt, U., Syperek, S., & Hertel, G. (2011). Testing on the Internet: Faking in a Web-Based Self-Administered Personality Measure. *Journal of Business and Media Psychology*, *1*, 1–11. Retrieved from www.journal-bmp.de

Laux, L., & Renner, K.-H. (2002). Self-Monitoring und Authentizität: Die verkannten Selbstdarsteller. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *23*(2), 129–148. https://doi.org/10.1024//0170-1789.23.2.129

Law, S. J., Bourdage, J., & O'Neill, T. A. (2016). To fake or not to fake: Antecedents to interview faking, warning instructions, and its impact on applicant reactions. *Frontiers in Psychology*, *7*, 1–13. https://doi.org/10.3389/fpsyg.2016.01771

LeBreton, J. M., Robin, J., Barksdale, C. D., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology*, *92*(1), 1–16. https://doi.org/10.1037/0021-9010.92.1.1

Lee, P., Joo, S. H., & Lee, S. (2019). Examining stability of personality profile solutions between Likert-type and multidimensional forced choice measure. *Personality and Individual Differences*, *142*(February), 13–20. https://doi.org/10.1016/j.paid.2019.01.022

Lee, H. S., Smith, W. Z., & Geisinger, K. F. (2017). Faking under a nonlinear relationship between personality assessment scores and job performance. *International Journal of Selection and Assessment*, *25*(3), 284–298. https://doi.org/10.1111/ijsa.12180

Lenhard, W. & Lenhard, A. (2016). *Berechnung von Effektstärken*. Abgerufen unter: https://www.psychometrica.de/effektstaerke.html. Psychometrica. DOI: 10.13140/RG.2.2.17823.92329

Lester, C., Anglim, J., & Fullarton, C. (2015). Individual Differences in Intention to Fake Job Interviews: Personality, Self-Monitoring, and the Theory of Planned Behaviour. *Australasian Journal of Organisational Psychology*, *8*, 1–11. https://doi.org/10.1017/orp.2015.7

Levashina, J., & Campion, M. A. (2006). A Model of Faking Likelihood in the Employment Interview. *International Journal of Selection and Assessment*, *14*(4), 1–30.

Li, H., Fan, J., Zhao, G., Wang, M., Zheng, L., Meng, H., … Lievens, F. (2021). The Role of Emotions as Mechanisms of Mid-Test Warning Messages during Personality Testing: A Field Experiment. *Journal of Applied Psychology*. https://doi.org/10.1037//0021-9010.87.6.1019

Lievens, F., De Corte, W., & Schollaert, E. (2008). A Closer Look at the Frame-of-Reference Effect in Personality Scale Scores and Validity. *Journal of Applied Psychology*, *93*(2), 268–279. https://doi.org/10.1037/0021-9010.93.2.268

Liu, J., & Zhang, J. (2020). An Item-Level Analysis for Detecting Faking on Personality Tests: Appropriateness of Ideal Point Item Response Theory Models. *Frontiers in Psychology*, *10*(January). https://doi.org/10.3389/fpsyg.2019.03090

Mammadov, S. (2022). Big Five personality traits and academic performance: A meta-analysis. *Journal of Personality*, *90*(2), 222–255. https://doi.org/10.1111/jopy.12663

Marasi, S., Wall, A., & Brewer, K. (2019). Participant carelessness: Is it a substantial problem with survey data? *Electronic Journal of Business Research Methods*, *17*(1), 1–16.

Marcus, B. (2009). “Faking” from the applicant’s perspective: A theory of self-presentation in personnel selection settings. *International Journal of Selection and Assessment*, *17*(4), 417–430. https://doi.org/10.1111/j.1468-2389.2009.00483.x

Martin, B. A., Bowen, C.-C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, *32*, 247–256. Retrieved from http://www.sciencedirect.com/science/article/pii/S0191886901000216%5Cnhttp://www.sciencedirect.com/science/article/pii/S0191886901000216/pdfft?md5=0a71450847312eaee2109f1f03efc3e7&pid=1-s2.0-S0191886901000216-main.pdf

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*(5), 812–821. https://doi.org/10.1037/0021-9010.85.5.812

McLarnon, M. J. W., DeLongchamp, A. C., & Schneider, T. J. (2019). Faking it! Individual differences in types and degrees of faking behavior. *Personality and Individual Differences*, *138*, 88–95. https://doi.org/10.1016/j.paid.2018.09.024

Melchers, K. G., Roulin, N., & Buehl, A. K. (2020). A review of applicant faking in selection interviews. *International Journal of Selection and Assessment*, *28*(2), 123–142. https://doi.org/10.1111/ijsa.12280

Mischel, W., & Shoda, Y. (1995). A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structire. *Psychological Review*, *102*(2), 246–268.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, *88*(2), 348–355. https://doi.org/10.1037/0021-9010.88.2.348

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017). Measuring non-cognitive predictors in high-stakes contexts: The effect of self-presentation on self-report instruments used in admission to higher education. *Personality and Individual Differences*, *106*, 183–189. https://doi.org/10.1016/j.paid.2016.11.014

Orne, M. T. (1962). On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and their Implications. *American Psychologist*, *17*(11), 776–783. Retrieved from http://search.proquest.com/docview/60621400?accountid=13042

Paulhus, D. L. (2011). Overclaiming on Personality Questionnaires. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New Perspectives on Faking in Personality Assessment*. https://doi.org/10.1093/acprof

König, C. J., Merz, A. S., & Trauffer, N. (2012). What is in Applicants' Minds When They Fill Out a Personality Test? Insights from a qualitative study. *International Journal of Selection and Assessment*, *20*(4), 442–452. https://doi.org/10.1111/ijsa.12007

Kunda, Z., & Sanitioso, R. (1989). Motivated changes in the self-concept. *Journal of Experimental Social Psychology*, *25*(3), 272–285. https://doi.org/10.1016/0022-1031(89)90023-1

R Core Team (2019). R: A language and environment for statistical computing (Version 3.6.0) [Computer software]. The R Foundation. https://www.R-project. org/.

Reeder, M. C., & Ryan, A. M. (2011). Methods for Correcting for Faking. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New Perspectives on Faking in Personality Assessment*. https://doi.org/10.1093/acprof

Robie, C., Born, M. P., & Schmit, M. J. (2001). Personal and situational determinants of personality responses: A partial reanalysis and reinterpretation of the Schmit et al. (1995) data. *Journal of Business and Psychology*, *16*(1), 101–117. https://doi.org/10.1023/A:1007843906550

Robie, C., Schmit, M. J., Ryan, A. M., & Zickar, M. J. (2000). Effects of Item Context Specificity on the Measurement Equivalence of a Personality Inventory. *Organizational Research Methods*, *3*(4), 348–365. https://doi.org/10.1177/109442810034003

Roess, M., & Roche, M. (2017). Job desire and response distortion in personality assessments. *E-Journal of Social & Behavioural Research in Business*, *8*(2), 1–15.

Röhner, J., Schröder-Abé, M., & Schütz, A. (2011). Exaggeration is harder than understatement, but practice makes perfect! faking success in the IAT. *Experimental Psychology*, *58*(6), 464–472. https://doi.org/10.1027/1618-3169/a000114

Röhner, J., & Schütz, A. (2020). Verfälschungsverhalten in psychologischer Diagnostik. *Report Psychologie*, *9*, 16–23.

Röhner, J., Thoss, P., & Schütz, A. (2022). Lying on the Dissection Table: Anatomizing Faked Responses. *Behavior Research Methods*, (December 2021). https://doi.org/10.3758/s13428-021-01770-8

Roulin, N., & Krings, F. (2019). Faking to Fit In: Applicants' Response Strategies to Match Organizational Culture. *Journal of Applied Psychology*. https://doi.org/10.1037/apl0000431

Roulin, N., Krings, F., & Binggeli, S. (2016). A dynamic model of applicant faking. *Organizational Psychology Review*, *6*(2), 145–170. https://doi.org/10.1177/2041386615580875

Salgado, J. F., & Lado, M. (2018). Faking resistance of a quasi-ipsative forced-choice personality inventory without algebraic dependence. *Journal of Work and Organizational Psychology*, *34*(3), 213–216. https://doi.org/10.5093/jwop2018a23

Schilling, M., Becker, N., Grabenhorst, M. M., & König, C. J. (2020). The relationship between cognitive ability and personality scores in selection situations: A meta-analysis. *International Journal of Selection and Assessment*, *29*(1), 1–18. https://doi.org/10.1111/ijsa.12314

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-Reference Effects on Personality Scale Scores and Criterion-Related Validity. *Journal of Applied Psychology*, *80*(5), 607–620. https://doi.org/10.1037/0021-9010.80.5.607

Schmitt, N., & Bradburn, J. C. (2018). An Ideal Student Factor and the validity of noncognitive measures of student potential. *International Journal of Selection and Assessment*, *26*(2–4), 109–123. https://doi.org/10.1111/ijsa.12221

Schulze, J., West, S. G., Freudenstein, J. P., Schäpers, P., Mussel, P., Eid, M., & Krumm, S. (2021). Hidden framings and hidden asymmetries in the measurement of personality—A combined lens-model and frame-of-reference perspective. *Journal of Personality*, *89*(2), 357–375. https://doi.org/10.1111/jopy.12586

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, *65*(3), 445–494. https://doi.org/10.1111/j.1744-6570.2012.01250.x

Sisco, H., & Reilly, R. R. (2010). Effect of Item Level Social Desirability on Factor Structure Stability. *International Journal of Selection and Assessment*, *18*(3), 305–309.

Stewart, G. L., Darnold, T. C., Zimmerman, R. D., Parks, L., & Dustin, S. L. (2010). Exploring how response distortion of personality measures affects individuals. *Personality and Individual Differences*, *49*(6), 622–628. https://doi.org/10.1016/j.paid.2010.05.035

Snyder, M., & Gangestad, S. (1986). On the Nature of Self-Monitoring. Matters of Assessment, Matters of Validity. *Journal of Personality and Social Psychology*, *51*(1), 125–139. https://doi.org/10.1037/0022-3514.51.1.125

Swift, V., & Peterson, J. B. (2019). Contextualization as a means to improve the predictive validity of personality models. *Personality and Individual Differences*, *144*(February 2018), 153–163. https://doi.org/10.1016/j.paid.2019.03.007

Terry, D. J., Hogg, M. A., & White, K. M. (1999). The theory of planned behaviour : Self-Identity, social identity and group norms. *British Journal of Social Psychology*, *38*(3), 225–244. https://doi.org/10.1348/014466699164149

Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E., & Coaster, J. (2012). Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability, and job type. *Personality and Individual Differences*, *52*(2), 195–201. https://doi.org/10.1016/j.paid.2011.10.017

Tonković, M., Jerneić, Ž., & Galić, Z. (2012). The effects of faking on the construct validity of personality questionnaires: A direct faking measure approach. *Psychological Topics*, *21*(3), 1–18.

van Hooft, E. A. J., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301–316. https://doi.org/10.1037/a0025711

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, *59*(2), 197–210. https://doi.org/10.1177/00131649921969802

Wang, Q., Liao, Y., & Burns, G. N. (2021). General, Work-Specific, and Work-Role Conscientiousness Measures in Predicting Work Criteria: A Comparative Perspective. *Applied Psychology*, *70*(1), 358–383. https://doi.org/10.1111/apps.12234

Wetzel, E., Frick, S., & Brown, A. (2020). Does multidimensional forced-choice prevent faking? Comparing the susceptibility of the multidimensional forced-choice format and the rating scale format to faking. *Psychological Assessment*. https://doi.org/10.1037/pas0000971

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, *7*(NOV), 1–12. https://doi.org/10.3389/fpsyg.2016.01832

Williams, M. M., Rogers, R., Sharf, A. J., & Ross, C. A. (2019). Faking Good: An Investigation of Social Desirability and Defensiveness in an Inpatient Sample With Personality Disorder Traits. *Journal of Personality Assessment*, *101*(3), 253–263. https://doi.org/10.1080/00223891.2018.1455691

Wright, J. C., & Mischel, W. (1987). A Conditional Approach to Dispositional Constructs: The Local Predictability of Social Behavior. *Journal of Personality and Social Psychology*, *53*(6), 1159–1177. https://doi.org/10.1037/0022-3514.53.6.1159

Wirtz, M. A. (2021). *Validität, interne im Dorsch Lexikon der Psychologie*. Retrieved from https://dorsch.hogrefe.com/stichwort/validitaet-interne

Ziegler, M., & Bensch, D. (2015). Personality Assessment, Faking and. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., Vol. 17, pp. 834–839). https://doi.org/10.1016/B978-0-08-097086-8.25083-6

Ziegler, M., MacCann, C., & Roberts, R. D. (2012). Faking. Knowns, Unknowns, and Points of Contention. In M. Ziegler, C. MacCann, & R. Roberts (Eds.), *New Perspectives on Faking in Personality Assessment* (pp. 3–16). https://doi.org/10.1093/acprof

Appendix

Informed Consent at Wave 1

Liebe Teilnehmer:innen,

vielen Dank, für Ihr Interesse an dieser Studie. Sie richtet sich an **Studierende aller Fachrichtungen**.

Bei dieser Untersuchung geht es um die Erprobung von Instruktionen im Rahmen der Testkonstruktion.

Die Studie wird **vier Erhebungszeitpunkte** haben, die im **Abstand von etwa 2 Wochen** stattfinden werden. Sie werden hierzu per Mail kontaktiert und haben jeweils **drei Tage** Zeit, den entsprechenden Fragebogen auszufüllen. Diese erste Befragung wird mit etwa 45 Minuten die längste sein. Die darauf folgenden Messzeitpunkte werden mit 10 bis 15 Minuten deutlich kürzer ausfallen.

Bitte versuchen Sie, die Befragungen in einer **ruhigen Umgebung** auszufüllen. Lesen und beantworten Sie die Fragen **sorgfältig**. Sie können so zu der Aussagekraft unserer Forschung beitragen.

Teilnahme am Gewinnspiel

Wenn Sie an allen vier Messzeitpunkten teilgenommen haben, werden Sie die Möglichkeit bekommen, an einem **Gewinnspiel** teilzunehmen. Verlost werden **Gutscheine im Wert von 1x50, 4x20 und 5x10 Euro für den *Avocadostore* oder für die Bamberger *CitySchexs****.
Studierenden der Universität Bamberg können außerdem **2 Versuchspersonenstunden** angerechnet werden.

Wenn Sie **Fragen oder Bedenken** bezüglich einer Teilnahme an dieser Studie haben, können Sie uns unter **greta-mia-louise.degro@stud.uni-bamberg.de** kontaktieren.

**Der Avocadostore ist ein Onlinehändler für nachhaltige Mode und fair gehandelte Produkte aus diversen Bereichen. Teilnehmende Bamberger:innen oder Personen, die planen, Bamberg einmal zu besuchen, können sich alternativ für Bamberger CitySchexs entscheiden und die teilnehmenden lokalen Unternehmen unterstützen.*

Datenschutz und Einwilligung

Die Richtlinien guter ethischer Forschung sehen vor, dass sich die Teilnehmer:innen an empirischen Studien explizit und nachvollziehbar mit der Teilnahme an der Studie einverstanden erklären.

Die Teilnahme an dieser Befragung erfolgt **freiwillig**. Sie können die Teilnahme jederzeit und ohne die Angabe von Gründen beenden.

Um Sie über die **weiteren Erhebungszeitpunkte** informieren zu können und für die **individuelle Rückmeldung** oder die **Teilnahme am Gewinnspiel** nach dem letzten Erhebungszeitpunkt, benötigen wir eine **E-Mail-Adresse** von Ihnen. Diese Adresse wird am Ende der Befragung erhoben und **getrennt** von den anderen Angaben, die Sie in der Befragung machen, **gespeichert**. So bleiben alle anderen in dieser Befragung erhobenen Daten **vollständig anonym**.

Mit der Durchführung der Befragung und Erhebung Ihrer Daten ist das Unternehmen SoSci Survey GmbH (www.soscisurvey.de/de/contact) gemäß Art. 28 DSGVO beauftragt. Die Datenschutzhinweise des Auftragsverarbeiters sind unter www.soscisurvey.de/de/data-protection einzusehen.

Hiermit bestätige ich, die Datenschutzinformationen gelesen zu haben und erkläre mich freiwillig zu einer Teilnahme an der Befragung bereit.

- ○ Ja, ich möchte an der Befragung teilnehmen.
- ○ Nein, ich möchte doch nicht an der Befragung teilnehmen.

Informed Consent at Wave 2

Willkommen zurück!

Liebe Teilnehmer:innen,
willkommen zum zweiten Teil der Befragung zur Erprobung von Instruktionen im Rahmen der Testkonstruktion.

Dieser Teil wird mit ca. 10 bis 15 Minuten deutlich kürzer ausfallen als die erste Erhebung.

Bitte versuchen Sie, die Befragungen in einer ruhigen Umgebung auszufüllen und lesen und beantworten Sie die Fragen sorgfältig. Sie können so zu der Aussagekraft unserer Forschung beitragen.

Gewinnspiel

Wenn Sie an allen vier Messzeitpunkten teilgenommen haben, werden Sie die Möglichkeit bekommen, an einem Gewinnspiel teilzunehmen. Verlost werden **Gutscheine im Wert von 1x50, 4x20 und 5x10 Euro für den *Avocadostore* oder für die Bamberger: *CitySchexs****.
Studierenden der Universität Bamberg können außerdem 2 Versuchspersonenstunden angerechnet werden.

Wenn Sie Fragen oder Bedenken bezüglich einer Teilnahme an dieser Studie haben, können Sie uns unter greta-mia-louise.degro@stud.uni-bamberg.de kontaktieren.

**Der Avocadostore ist ein Onlinehändler für nachhaltige Mode und fair gehandelte Produkte aus diversen Bereichen. Teilnehmende Bamberger:innen oder Personen, die planen, Bamberg einmal zu besuchen, können sich alternativ für Bamberger CitySchexs entscheiden und die teilnehmenden lokalen Unternehmen unterstützen.*

Hiermit erkläre ich mich freiwillig zu einer Teilnahme an diesem Teil der Befragung bereit.

- ◯ Ja, ich möchte an der Befragung teilnehmen.
- ◯ Nein, ich möchte doch nicht an der Befragung teilnehmen.

Informed Consent at Wave 3

Willkommen zurück!

Liebe Teilnehmer:innen,
willkommen zum dritten Teil der Befragung zur Erprobung von Instruktionen im Rahmen der Testkonstruktion.

Dieser Teil wird ähnlich wie der zweite Teil etwa 10-15 Minuten dauern.

Bitte versuchen Sie, die Befragungen in einer ruhigen Umgebung auszufüllen und lesen und beantworten Sie die Fragen sorgfältig. Sie können so zu der Aussagekraft unserer Forschung beitragen.

Gewinnspiel

Wenn Sie an allen vier Messzeitpunkten teilgenommen haben, werden Sie die Möglichkeit bekommen, an einem Gewinnspiel teilzunehmen. Verlost werden **Gutscheine im Wert von 1x50, 4x20 und 5x10 Euro für den *Avocadostore* oder für die Bamberger: *CitySchexs****.
Studierenden der Universität Bamberg können außerdem 2 Versuchspersonenstunden angerechnet werden.

Wenn Sie Fragen oder Bedenken bezüglich einer Teilnahme an dieser Studie haben, können Sie uns unter greta-mia-louise.degro@stud.uni-bamberg.de kontaktieren.

**Der Avocadostore ist ein Onlinehändler für nachhaltige Mode und fair gehandelte Produkte aus diversen Bereichen. Teilnehmende Bamberger:innen oder Personen, die planen, Bamberg einmal zu besuchen, können sich alternativ für Bamberger CitySchexs entscheiden und die teilnehmenden lokalen Unternehmen unterstützen.*

Hiermit erkläre ich mich freiwillig zu einer Teilnahme an diesem Teil der Befragung bereit.

- ○ Ja, ich möchte an der Befragung teilnehmen.
- ○ Nein, ich möchte doch nicht an der Befragung teilnehmen.

Informed Consent at Wave 4

Willkommen zurück!

Liebe Teilnehmer:innen,
willkommen zum vierten und letzten Teil der Befragung zur Erprobung von Instruktionen im Rahmen der Testkonstruktion.

Dieser Teil wird etwa 15 Minuten dauern.

Bitte versuchen Sie, die Befragungen in einer ruhigen Umgebung auszufüllen und lesen und beantworten Sie die Fragen sorgfältig. Sie können so zu der Aussagekraft unserer Forschung beitragen.

Gewinnspiel

Im Anschluss an diese Befragung, werden Sie die Möglichkeit bekommen, an einem Gewinnspiel teilzunehmen. Verlost werden **Gutscheine im Wert von 1x50, 4x20 und 5x10 Euro für den *Avocadostore* oder für die Bamberger: *CitySchexs****.
Studierenden der Universität Bamberg können außerdem 2 Versuchspersonenstunden angerechnet werden.

Wenn Sie Fragen oder Bedenken bezüglich einer Teilnahme an dieser Studie haben, können Sie uns unter greta-mia-louise.degro@stud.uni-bamberg.de kontaktieren.

**Der Avocadostore ist ein Onlinehändler für nachhaltige Mode und fair gehandelte Produkte aus diversen Bereichen. Teilnehmende Bamberger:innen oder Personen, die planen, Bamberg einmal zu besuchen, können sich alternativ für Bamberger CitySchexs entscheiden und die teilnehmenden lokalen Unternehmen unterstützen.*

Hiermit erkläre ich mich freiwillig zu einer Teilnahme an diesem Teil der Befragung bereit.

- ○ Ja, ich möchte an der Befragung teilnehmen.
- ○ Nein, ich möchte doch nicht an der Befragung teilnehmen.

Participant-Code and Demographic Variables

Participant-Code
Damit wir Ihre Antworten aus den vier Messzeitpunkten einander zuordnen und Sie trotzdem anonym bleiben können, bitten wir Sie, sich nun selbst Ihren Versuchspersonen-Code zu erstellen.
Tragen Sie hierzu die entsprechenden Buchstaben und Zahlen in das Feld unten ein.

Sie müssen sich den Code nicht merken. Die Anweisungen zum Erstellen des Codes werden zu jedem Messzeitpunkt vollständig präsentiert.

1. Der zweite Buchstabe Ihres Nachnamens:
2. Die letzten beiden Ziffern Ihres Geburtsjahres:
3. Der erste Buchstabe des Vornamens Ihres Vaters:
4. Der erste Buchstabe des Vornamens Ihrer Mutter:
5. Ihre Hausnummer (dreistellig, ohne Buchstaben)

Beispiel: Frau Schmidt (S) ist im Jahr 1973 (73) geboren. Ihr Vater heißt Norbert (N) und ihre Mutter Dagmar (D). Ihre Hausnummer ist die 4B. (004)
Der Versuchspersonencode würde heißen: S73ND004

Bitte erstellen Sie nun Ihren Versuchspersonencode und tragen Sie ihn in das freie Feld ein.

Demographic Data
Nun interessieren wir uns für einige demografische Daten von Ihnen.

- Welchem Geschlecht ordnen Sie sich am ehesten zu?
  Männlich/ Weiblich/ Divers
- Wie lautet Ihr Geburtsdatum?
  TT.MM.JJJ
- An welcher Universität studieren Sie? *Bitte keine Abkürzungen, sondern den vollständigen Namen der Universität eintragen.*
- Wie lautet die offizielle Bezeichnung Ihres Studiengangs?
- Wie lautet Ihre aktuelle Durchschnittsnote?

Conscientiousness Measures: Instruction Types and Item Wording

Generic/ „Honest“ Instructions

Nachstehend finden Sie eine Reihe von Eigenschaften, die auf Sie zutreffen könnten. **Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.**
Es gibt dabei keine richtigen oder falschen Antworten. Wählen Sie immer diejenige Antwort aus, die Sie am besten beschreibt. **Seien Sie dabei bitte so ehrlich wie möglich.**

Applicant/ “Faking” Instructions

Sie bewerben sich für einen sehr beliebten und gut bezahlten Hiwi-Job mit langfristiger Perspektive. Nachstehend finden Sie eine Reihe von Eigenschaften, die in diesem Job auf Sie zutreffen könnten. **Bitte geben Sie für jede der folgenden Aussagen an, inwieweit Sie zustimmen.**
Beantworten Sie den folgenden Fragebogen so, dass Sie sich von Ihrer besten Seite zeigen, um so Ihre Chancen zu erhöhen, den Job zu bekommen.

Non-contextualized Items

- Ich bin eher unordentlich.
- Ich bin bequem, neige zu Faulheit.
- Ich bin stetig, beständig.
- Ich bin systematisch, halte meine Sachen in Ordnung.
- Ich neige dazu, Aufgaben vor mir herzuschieben.
- Ich bin manchmal ziemlich nachlässig.
- Ich mag es sauber und aufgeräumt.
- Ich bin effizient, erledige Dinge schnell.
- Ich fühle mich oft bedrückt, freudlos.
- Ich bin verlässlich, auf mich kann man zählen.
- Ich bin eher der chaotische Typ, mache selten sauber.
- Ich bleibe an einer Aufgabe dran, bis sie erledigt ist.
- Manchmal verhalte ich mich verantwortungslos, leichtsinnig.

Contextualized Items

- Ich bin im Arbeitskontext eher unordentlich.
- Ich bin im Arbeitskontext bequem, neige zu Faulheit.
- Ich bin im Arbeitskontext stetig, beständig.
- Ich bin im Arbeitskontext systematisch, halte meine Sachen in Ordnung.
- Ich neige dazu, Aufgaben vor mir herzuschieben.
- Ich bin im Arbeitskontext manchmal ziemlich nachlässig.
- Ich bin im Arbeitskontext effizient, erledige Dinge schnell.
- Ich fühle mich im Arbeitskontext oft bedrückt, freudlos.
- Ich bin im Arbeitskontext verlässlich, auf mich kann man zählen.
- Ich bin im Arbeitskontext eher der chaotische Typ, mache selten sauber.
- Ich bleibe im Arbeitskontext an einer Aufgabe dran, bis sie erledigt ist.
- Manchmal verhalte ich mich im Arbeitskontext verantwortungslos, leichtsinnig.

Scale

- Stimme überhaupt nicht zu.
- Stimme eher nicht zu.
- Teils, teils.
- Stimme eher zu.
- Stimme voll und ganz zu.

Adopted Frage of Reference

Bei der Beantwortung dieser Frage habe ich …

- an mein typisches Verhalten im Bildungskontext gedacht (z.B. in der Schule, in der Uni oder in Lerngruppen)
- an mein typisches Verhalten im Arbeitskontext gedacht (z.B. wenn ich arbeitsbezogene Aufgaben erledige, wie ich meinen Arbeitsplatz gestalte, wie ich mit Kolleg:innen oder Vorgesetzten während der Arbeitszeit interagiere).
- an mein typisches Verhalten in sozialen Kontexten gedacht (z.B. in der Interaktion mit meiner Familie oder Freunden, in meiner Freizeit).
- an mein typisches Verhalten in unterschiedlichen Kontexten gedacht (also z.B. sowohl im Arbeitskontext als auch in sozialen Kontexten).
- an einen anderen Kontext gedacht, und zwar:

Scale

Ja
Nein

Motivation and Environment

Wir würden nun noch gerne etwas über Ihre Motivation und Ihre Umgebung während der Befragung erfahren.

- Ich habe sorgfältig geantwortet.
- Ich konnte mich gut in die Situation hineinversetzen. (*only for G+ and A+ conditions)*
- Ich habe mir für die Beantwortung Zeit genommen.
- Ich habe mich bei der Beantwortung der Fragen beeilt.
- Ich habe den Fragebogen in einer ruhigen Umgebung ausgefüllt.

Scale

- Stimme überhaupt nicht zu.
- Stimme eher nicht zu.
- Teils, teils.
- Stimme eher zu.
- Stimme voll und ganz zu.

Self-Monitoring

Wir möchten nun noch etwas über Ihr Verhalten im Allgemeinen erfahren. Inwieweit stimmen Sie den unten stehenden Aussagen zu? *Wählen Sie immer diejenige Auswahlmöglichkeit, die am besten zu Ihnen passt. Es gibt dabei keine richtigen oder falschen Antworten.*

- Es ist wichtig für mich, mich in die Gruppe, in der ich mich gerade aufhalte, einzupassen.
- Um nicht aufzufallen, verhalte ich mich in verschiedenen Situationen und mit verschiedenen Leuten oft wie ganz unterschiedliche Personen.
- Wenn ich in sozialen Situationen den Eindruck habe, dass eine andere Verhaltensweise angebracht ist, dann kann ich mein Verhalten problemlos umstellen.
- Der geringste Hinweis von Missbilligung in den Augen einer anderen Person genügt, damit ich mein Verhalten ändere.
- Ich versuche die Reaktionen anderer auf mein Verhalten zu registrieren, damit ich mich nicht selbst ins Abseits stelle.
- Ich kann normalerweise an den Augen meines Gesprächspartners ablesen, ob ich etwas Unangemessenes gesagt habe.
- Wenn sich alle Personen in einer Gruppe auf eine bestimmte Art und Weise verhalten, dann habe ich das Gefühl, dass das die richtige Art, sich zu verhalten, sein muss.
- Normalerweise merke ich den Leuten an, dass sie einen Witz geschmacklos finden, selbst wenn sie darüber lachen.
- In Unterhaltungen registriere ich bereits die kleinste Veränderung im Gesichtsausdruck meines Gesprächspartners.
- Oftmals wage ich es nicht, mich gegenüber anderen Personen so zu verhalten, wie ich wirklich bin.
- Ich bin nicht immer die Person, die ich vorgebe zu sein.
- Ich habe manchmal das Gefühl, dass die Leute nicht wissen, wer ich wirklich bin.
- Ich kann mich ziemlich gut auf meine Intuition verlassen, wenn es darum geht, die Gefühle und Motive anderer zu verstehen.
- Wählen Sie hier bitte stimme eher nicht zu aus.
- Wenn ich nicht weiß, wie ich mich in einer bestimmten Situation verhalten soll, orientiere ich mich am Verhalten anderer.
- Um Missbilligung zu vermeiden, stelle ich gegenüber verschiedenen Leuten ganz unterschiedliche Aspekte meiner Persönlichkeit dar.
- Ich kann oft die wahren Gefühle einer Person an ihren Augen ablesen.
- Wenn mir jemand eine Lüge erzählt, dann merke ich das sofort an seiner/ihrer Ausdrucksweise.
- Wenn ich das Gefühl habe, dass das Bild meiner Persönlichkeit, welches ich in sozialen Situationen vermitteln will, nicht ankommt, dann kann ich es jederzeit in ein angemesseneres umwandeln.
- Es fällt mir schwer mein Verhalten unterschiedlichen Situationen und Leuten anzupassen.
- Ich habe die Erfahrung gemacht, dass ich mein Verhalten so steuern kann, dass ich den Anforderungen jeder Situation gerecht werden kann.
- Verschiedene Situationen bringen mich dazu, mich wie sehr unterschiedliche Personen zu verhalten.
- Wenn ich einen ganz bestimmten Eindruck von mir vermitteln will, dann kann ich das auch entsprechend steuern.
- Mein Verhalten ist oft so, wie andere es sich wünschen.
- Sobald ich weiß, welches Verhalten eine bestimmte Situation erfordert, kann ich mich problemlos darauf einstellen.

Scale

- Stimme überhaupt nicht zu.
- Stimme eher nicht zu.
- Teils, teils.
- Stimme eher zu.
- Stimme voll und ganz zu.

Experience and Knowledge

Nun möchten wir erfahren, ob Sie bereits Berührungspunkte mit der Tätigkeit als studentische Hilfskraft hatten und wie Sie über einen solchen Job denken.
Bitte kreuzen Sie diejenige Auswahlmöglichkeit an, die für Sie am besten passt. Es gibt dabei keine richtigen oder falschen Antworten.

- Ich habe mich selbst schon einmal als studentische Hilfskraft beworben.
- Ich bin bzw. war studentische Hilfskraft.
- Ich habe Bekannte, die als studentische Hilfskraft arbeiten bzw. gearbeitet haben.
- Ich habe davon gehört, oder darüber gelesen und glaube, mir gut vorstellen zu können, was es bedeutet studentische Hilfskraft zu sein.
- Ich habe mich noch nicht mit der Bewerbung für einen Job als studentische Hilfskraft beschäftigt.
- Ich kann mir nicht viel unter einem Job als studentische Hilfskraft vorstellen.
- Ich lehne einen Job als studentische Hilfskraft für mich ab.
- Ich interessiere mich dafür, wie es ist, als studentische Hilfskraft zu arbeiten.
- Ich möchte mich damit nicht beschäftigen, was eine studentische Hilfskraft tut.
- Ich finde, ein Job als studentische Hilfskraft passt nicht zu mir.

Scale

- Stimme überhaupt nicht zu.
- Stimme eher nicht zu.
- Teils, teils.
- Stimme eher zu.
- Stimme voll und ganz zu.

Assumptions Regarding the Purpose of the Study

Was glauben Sie, sollte mit der Studie untersucht werden?
Bitte schreiben Sie Ihre Ideen in das freie Feld unten. Uns interessieren dabei alle Ideen, die Sie zum Ziel der Studie haben. Es gibt dabei keine richtigen oder falschen Antworten.

Collection of E-Mail-Adresses Wave 1 to 3

Vielen Dank für Ihre Teilnahme!

Sie haben nun das Ende der aktuellen Befragungswelle erreicht.

In etwa zwei Wochen erhalten Sie dann den Link zur nächsten Befragung.

Um Sie über den nächsten Erhebungszeitpunkt zu informieren und Ihnen den Link zum Fragebogen zuzuschicken, geben Sie bitte hier eine gültige E-Mail-Adresse an, deren Posteingang Sie regelmäßig prüfen.

Diese Adresse wird getrennt von Ihren anderen Antworten im gespeichert, sodass die Anonymität Ihrer Angaben gewahrt bleibt. Geben Sie bitte immer dieselbe E-Mail-Adresse an.

Sie können in etwa zwei Wochen mit der Benachrichtigung über die nächste Befragung rechnen.

E-Mail-Adresse:

Lottery at Wave 4

Vielen Dank für Ihre Teilnahme!

Sie haben nun das Ende der letzten Befragungswelle erreicht.

Wir möchten uns sehr herzlich für Ihre Unterstützung bedanken!

Vielen Dank, dass Sie an allen vier Erhebungszeitpunkten teilgenommen haben.

Sie haben nun die Möglichkeit, an einem Gewinnspiel teilzunehmen *oder* sich für eine individuelle Rückmeldung bezüglich Ihres Persönlichkeitsprofils einzutragen.

Studierende der Universität Bamberg können zudem 2 Versuchspersonenstunden angerechnet bekommen. Wählen Sie hierzu *zusätzlich* die letzte Antwortalternative aus.

Bitte wählen Sie höchstens **eine** Antwortoption aus.

Mit der Angabe Ihrer E-Mail-Adresse willigen Sie ein, dass Ihre E-Mail-Adresse bis zur Übermittlung der Rückmeldung gespeichert wird. Diese Einwilligung können Sie jederzeit widerrufen. Ihre Angaben in dieser Befragung bleiben weiterhin anonym, Ihre E-Mail-Adresse wird nicht an Dritte weitergegeben.

- ☐ Ich will am **Gewinnspiel** teilnehmen. Ich willige ein, dass meine E-Mail-Adresse bis zur Ziehung der Gewinner gespeichert wird. Diese Einwilligung kann ich jederzeit widerrufen. Meine Angaben in dieser Befragung bleiben weiterhin anonym, meine E-Mail-Adresse wird nicht an Dritte weitergegeben.
- ☐ Ich möchte eine **individuelle Rückmeldung** über mein Persönlichkeitsprofil erhalten.
- ☐ Ich studiere an der Uni Bamberg und möchte Versuchspersonenstunden erhalten. (Diese Option kann zusätzlichzur Teilnahme am Gewinnspiel ausgewählt werden.)
- ☐ Ich möchte nach Abschluss der Studie Informationen über die Inhalte und Ergebnisse der Studie erhalten. (Diese Option kann zusätzlichzur Teilnahme am Gewinnspiel ausgewählt werden.)